DOCUMENT RESUME

ED 427 575 HE 031 766

AUTHOR	Ridley, Dennis R.; Quanty, Michael B.; Sciabica, Mark
TITLE	Grading Standards and Course Challenge: An Analytical-Empirical Approach. ASHE Annual Meeting Paper.
PUB DATE	1998-11-06
NOTE	35p.; Paper presented at the Annual Meeting of the Association for the Study of Higher Education (23rd, Miami, FL, November 5-8, 1998).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	College Faculty; College Students; *Courses; Departments; *Difficulty Level; Educational Policy; *Grade Inflation; Grade Point Average; *Grading; Higher Education; *Student Attitudes; Student Evaluation of Teacher Performance; Universities
IDENTIFIERS	*ASHE Annual Meeting

ABSTRACT

This study examined whether conditions affecting intellectual challenge and course difficulty experienced by students were associated with the severity of grading standards. A policy by a university academic unit that reduced the number of high grades awarded by one of its departments provided an opportunity to test a research-based prediction related to this issue. It was predicted that mean grades would fall while courses would deliver greater intellectual challenge to students. Data were obtained from university records of grade distributions and surveys of student evaluations of instruction related to course difficulty and course challenge. It was found that after the policy reducing the number of high grades awarded to students was implemented, mean grades fell significantly while ratings of course challenge and difficulty rose significantly in relation to other courses. The results supported the hypothesis that monitoring grades given by faculty members can create changes that increase both the perceived difficulty and the challenge of the course. (Contains 18 references, notes and 3 tables.) (MDM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 427 575

Grading Standards and Course Challenge: An Analytical-Empirical Approach[1]

Dennis R. Ridley, Ph.D.
Christopher Newport University
Michael B. Quanty, Ph.D.
Thomas Nelson Community College
Mark Sciabica
Christopher Newport University

BEST COPY AVAILABLE

HF6131 766

Paper presented at the meetings of the Association for the Study of Higher Education, November 6, 1998.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
V. Vaughan
ASHE
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ASHE *Association for the Study of Higher Education*

Headquartered at the University of Missouri-Columbia • College of Education • Department of Educational Leadership and Policy Analysis • 211 Hill Hall, Columbia, MO 65211 • 573-882-9645 • fax 573-884-5714

This paper was presented at the annual meeting of the Association for the Study of Higher Education held in Miami, Florida, November 5-8, 1998. This paper was reviewed by ASHE and was judged to be of high quality and of interest to others concerned with higher education. It has therefore been selected to be included in the ERIC collection of ASHE conference papers.

Grading Standards and Course Challenge

An Analytical-Empirical Approach

Abstract

In this article, the authors explored the issue of whether the conditions affecting intellectual challenge and course difficulty experienced by students are frequently associated with the severity of grading standards. One university academic unit's policy reducing the number of high grades awarded by one of its departments provided a rare opportunity to test a research-based prediction related to this issue. It was predicted that mean grades would fall while courses delivered greater intellectual challenge. Grade distribution data, and student survey evaluations of instruction related to course difficulty and course challenge, provided the essential data. The data from courses offered inside and outside the department and before and after the policy change were analyzed. As predicted, mean grades fell significantly, while ratings of course challenge and difficulty rose significantly, relative to other courses. Important issues regarding the imposition of grade standards are discussed.

There is an ambiguity in the notions of "difficulty" and "challenge" as commonly applied to college courses. These descriptors can refer to different aspects of the college course experience. The American Heritage Dictionary (1985) includes the following definition of "challenge": "the quality of requiring full use of one's abilities, energy, or resources" (p. 256) whereas "difficult" is "hard to comprehend or solve" (p. 395). These would seem to be complementary definitions of course difficulty in that they require students to exert effort to demonstrate course mastery. There is, however, another sense in which a course can be defined as difficult: that is the likelihood of achieving a certain grade level.

The general issue studied in this article is whether the conditions of a course's difficulty that pertain to student effort frequently go together with the latter meaning, i.e., severity of grading standards. The counterexamples are not hard to find, such as, hypothetically, when a student's test or paper might be graded by two different instructors and given different grades. Such a grading exercise would change neither the difficulty of the subject nor the challenge to students since the same level of mastery translated into different grades given by two professors. In general, grades can be dissociated from the measurement

of educational benefits connected with a course (Basinger, 1997).

Nevertheless, an empirical test must decide whether the two conditions named above frequently go together. Let us call the position that they do go together a "latent trait" theory. According to this view, high grading standards are one expression of a latent trait such as scholarly rigor. By assumption, professors who vigilantly maintain high standards exercise an attitude that has other consequences, i.e., providing intellectual stimulation and challenge. One behavioral prediction would follow: if grading standards are raised, professors' sense of intellectual rigor will heighten, overflowing in a more challenging educational experience for students. This prediction is consistent with much social psychological research which suggests that attitude change consistent with a new policy or position will follow an induced behavior change in support of that position provided that the inducement is just sufficient to produce behavioral commitment (Cialdini, 1993).

The behavioral prediction and underlying theory are rarely stated as starkly as above. Nevertheless, such statements are consistent with much serious discussion linking grade inflation with the decline of academic standards. Rotfeld (1997) cited cases of

administrative pressure on teachers to increase the number of high grades awarded, thus maintaining enrollments at the cost of lowered academic standards. He then stated:

> ...if the university is really concerned about standards, if it really wants to make certain that graduation is a sign of intellectual development, it should focus on fighting grade inflation and supporting faculty against pressures to lower academic standards in the classroom. To raise standards, administrators and tenure committees should exhibit skepticism of teachers who repeatedly award almost everyone A grades every term... (Rotfeld, 1997, p. 9)

Rotfeld here came close to the above behavioral prediction, recommending limits on the number of high grades as a direct route to raising academic standards. His underlying theory perhaps differed, but that matters little until the behavioral prediction has been tested.

Immediate Study Background

One university academic unit's policy change made in 1993 provided a rare natural opportunity to test the above prediction. A formal program review of one academic department[2] included a recommendation that "Program faculty identify ways in which it might

further challenge students, thereby lowering the modal grade from 'A' to at most 'B.'" Further, the review stated, "Evaluation criteria must be raised in order to provide greater challenge, and as a by-product, to bring grades more in line with those awarded by the University as a whole." The administration responded to this recommendation by requiring that, for the years 1993-95, any department member assigning more than 50 percent A's in any course must provide written justification prior to the submission of grades.

Both the program reviewers and the administration clearly believed that grading standards and academic challenge went hand in hand. The program reviewers cited only grade distribution data as evidence of students' lack of challenge. This practice begged the question of how independent grading severity and challenge are. As argued earlier, grading severity and student challenge are distinct and one condition might exist without the other. Thus the researchers decided to find independent indicators of the level of challenge that students experienced.

The policy lasted officially as planned from 1993 to 1995. Further, a de facto policy remained through spring, 1996, the period corresponding to the key administrator's service. Program faculty raised the issue of how effective the policy had been. In

preparation for the next program review the faculty sought to document what happened in response to each recommendation and what the long-term effects had been.

Hypotheses

There were two hypotheses. Relative to the rest of the university: (1) the policy lowered mean grades in undergraduate courses; (2) the policy increased both the perceived difficulty and the reported intellectual challenge of the course or subject matter.

Methods

Data Sources

Undergraduate course grades were a main source of data. These grades were divided into those from the department under study and all others. Total department enrollments and course sections (excluding tutorials, practica, etc.) varied from semester to semester. They were in the approximate range of 600 to 850 enrollments in 25 to 35 courses per semester. The institution-wide numbers of each were approximately 20 times greater.

Concerning grade data, both aggregate and individual student data (only for courses in the discipline) were examined for the study institution. In addition to these data, substantial aggregate data provided by a sister institution offered a point of comparison.

The Instruction Evaluation Survey (IES) provided another main source of data. Primary interest focused on department-sponsored courses for undergraduates in comparison with courses outside the department. Most undergraduate classes routinely completed this survey during either the thirteenth or fourteenth week in the semester. Thus, the range of course sections represented was approximately as cited above. The number of respondents fell somewhat short of the above because it depended upon attendance when surveys were administered.

Procedure

A strategy to test the hypothesized effects derived from the student course evaluation data for each course. The student survey included a question asking students how difficult they found the subject matter of the course. Students responded to the statement, "The subject matter of this course is difficult" using a five-point Likert scale (5 = "Strongly Agree" to 1 = "Strongly Disagree"). Another item stated, "The instructor was intellectually motivating and stimulated learning." These two questions operationally defined, respectively, "difficulty" and "course challenge." In addition, actual grade data were examined for both departmental courses and the institution as a whole. A related item

on the student evaluation survey asked the students' expected grade in the course (A-F and "Other").

Two other questions from the student course evaluation survey also were examined. The first was, "The instructor's grading procedures were fair." The second stated, "Tests covered knowledge, application, or reasoning that could be expected on the basis of course content." These items were chosen as still other indicators of grading severity not necessarily related to challenging students.

The researchers compared the two-year periods (1991-93 and 1994-96) which immediately preceded and followed the policy. One year (1993-94) between these periods was the first year of the policy. This period provided an opportunity for the policy to become established and possibly influence student perceptions of the courses.

Separate analyses of variance were carried out for each variable. Each analysis tested the effects of two variables, the discipline under study versus all other enrollments and the time dimension (pre- vs. post-policy), plus the interactions between these two variables. Interactions, if significant, would suggest change associated with the policy. Each interaction would show a difference between pre- and post-policy periods that depended on the discipline. Three survey

variables confined the pre-policy data tc the last year (1992-93); earlier data were lost due to archive retirement. Repeated measures were not applied because the survey data records preserved the key information for each variable but not individual students' records on all variables.

Supplementary analyses were also done. First, a one-time university-wide change to a plus and minus grading scale occurred concomitantly with the administrative review studied in this article. This created the opportunity for a further analysis of how any changes in mean grades corresponded with the use of plus and minus grade options. A second supplementary analysis examined more recent data at the study institution. These data included grades and student survey data for the three most recent semesters. Another supplementary analysis was done by creating a "control group" for comparison purposes. The comparison examined grades at a neighboring university for the same time period and applied similar analyses to the grades awarded inside and outside of the corresponding department.

Finally, a colleague suggested a crucial supplementary analysis to test the policy effects, if any, for different general academic achievement levels.[3] For example, one would expect a limitation of

A-grades to have little effect on the g.p.a.'s of students at either the upper or lower g.p.a. ranges. High achieving students may continue receiving A's; students who almost never earn A's also would be unaffected. Further, the cohorts taking courses in the study discipline in the pre-policy and post-policy periods were largely separate and distinct. Therefore, it is important to test whether apparent effects on grades (by extension, on course challenge) were due to a general decline in academic ability. To test these notions, the research selected a random sample of each cohort (i.e., before and after the policy), insuring that each member of the sample had taken at least two courses within the discipline and two courses from outside during the same period of time. In-department and outside-department g.p.a.'s for both cohorts provided the crucial data for comparison. An analysis of variance was done for the sample as described. In addition, the analysis was repeated for the middle 75 percent of students in each group arranged by overall g.p.a., based on the assumptions that the highest and lowest achieving students will be least affected by a policy controlling A-grades.

Results

Analyses of variance were done on the following six variables: actual grades and expected grades at the

study institution, subject matter difficulty, stimulation of learning, grading fairness, and content appropriateness of tests.[4] There were two discipline levels: inside vs. outside the study discipline and two levels of time: 1991-93 vs. 1994-96.

Hypothesis 1

Analyses of grades and expected grades pertained to the first hypothesis. The two grade analyses of variance both showed a significant interaction as predicted. Actual mean grades stayed about the same for enrollments outside the department (-.01 grade points) but lowered significantly for department enrollments (-.14 grade points). Expected grades also showed significant changes: they tended to go up for the outside-department enrollments (+.11 grade points) while declining for department enrollments (-.22 grade points).

Insert Table 1 about here

Supplementary analyses revealed, first, a significant drop in A grades among department course grades after the policy took effect, even when adding together A's and the new A-minuses in the post-policy period. However, adding B-pluses into the high grade category in the later period resulted in a pattern of

only 1.5 percent more grades in that category--a non-significant difference. In contrast, non-department grade data revealed that the combined A and A- were used with almost exactly the same frequency as the former A grade alone. Another supplementary analysis of the three most recent semesters also revealed that discipline course grades remained at the post-treatment level.

Insert Table 2 about here

A supplementary analysis of grade data from a neighboring university showed a similar pattern. Grades in the department (which corresponded to that targeted by the policy in the study university) were considerably higher than that university's average in the 1991-93 period and dropped significantly during the 1994-96 period. Although the department's mean grades remained significantly higher than the university's, the same interaction occurred; i.e., the mean grades dropped in the department's courses but not generally in other departments' courses.

Insert Table 3 about here

Two further comparisons tested the limits of

similarity between the two institutions' departmental grade data. The first compared the grade distributions on the frequencies of A grades (including A- throughout this paragraph) versus all others. Comparing the semesters immediately before and after the change, the neighboring university awarded 3.9 percent fewer A's and the study university awarded 11.4 percent fewer A's. The second comparison focused on the timing of changes. The neighboring university showed a drop of 3.6 percent A's during the year prior to the policy (i.e. almost equal to that from before to after the policy period). The study university showed an increase of 0.9 percent A's during that period. (The comparison for periods bracketing the policy change is reported above.)

A final supplementary analysis tested whether grade changes reflected a general decline in academic achievement of the two cohorts before and after the policy. The researchers selected a random sample of 160 students enrolled in at least two discipline and two non-discipline courses during the pre-policy period (N = 80) or post-policy period (N = 80). Each student's mean grades within and outside the discipline comprised the key data. An analysis of variance tested the effects of time, discipline, and their interaction on g.p.a.'s.

The analysis for the unrestricted sample showed the familiar highly significant effect of discipline. G.p.a.'s were higher in the discipline in both cohorts. There were no other significant effects. However, repeating the analysis for the middle 75 percent of cases based on overall g.p.a. revealed that all three effects (discipline, time, and interaction) were significant. Comparing the two cohorts, there was a general decline in g.p.a.'s for the combined discipline and non-discipline courses of about .22 grade points. However, g.p.a.'s for discipline courses declined approximately .35 grade points versus .10 for non-discipline courses.

Hypothesis 2

The next two analyses of variance pertained to the second hypothesis relating to course difficulty and challenge. They were based on the items: "The subject matter of this course is difficult," and "The instructor was intellectually motivating and stimulated learning." These two items also produced similar results. Department students significantly increased their ratings between the pre- and post- policy periods; that is, they reported courses as more difficult and instructors more stimulating. Non-department students revealed a different pattern of results. For the first item, course difficulty, in the

post-policy period department enrollees revealed significantly higher difficulty ratings by a mean of .24 on a 5-point scale. Other enrollees showed a decline in difficulty ratings by a mean of .14. Thus, there was a net difference of .38 between department and non-department enrollees on this variable. For the second item, stimulation of learning, both department and other students gave significantly higher mean ratings after the policy, but the department students did so to a much greater degree. These two measures of increase diverged by .34. Thus, after the policy, students enrolled in departmental courses rated these courses both more difficult and more challenging than before the policy to a greater extent than students enrolled in other disciplines' courses.[5]

Two other variables, grading fairness and test content appropriateness, also had a similar pattern of results. These items were: "The instructor's grading procedures were fair," and "Tests covered knowledge, application, or reasoning that could be expected on the basis of course content." In both cases, department students gave slightly higher ratings (again significant) in the later period, but outside students tended to do the same. Another way these two variables differed was that department students' ratings tended to be lower than other students'. However, there was

no significant interaction effect in either case.

Supplementary analysis of end-of-course student ratings also revealed that student's ratings of course difficulty remained at the post-treatment level. Students' ratings of how stimulating instructors were fell significantly in the department relative to other departments, but they remained significantly above the pre-treatment base line.

In sum, the two time periods saw two general changes that might be associated with the policy. First, grades lowered significantly for the department's courses after its policy went into effect. This change was not associated with a change in the general academic achievement of the cohorts enrolled in these courses. Second, other results suggest that the policy change had the desired effects. Students perceived department courses as more difficult *and* instructors as more intellectually motivating and stimulating. At the same time, perceptions of fairness in grading and appropriateness of test content did not change relative to the rest of the University.

Discussion

In view of its intention to lower mean grades as observed, the grading policy was one important contributor to that outcome. Both administrative monitoring of grading and self-monitoring by professors

apparently decreased the number of A's and thus depressed mean grades. While other changes occurred in the department during the period in question, none aimed directly at changing grades as the policy had done.

Instructors might have achieved compliance with the grading policy merely by changing the level of mastery required for an A. However, departmental students did not find their evaluations inappropriate or arbitrary. Arguably, grading practices went along with more fundamental changes in standards or expectations.

The study supported the hypothesis that monitoring faculty members' grades can create changes that increase both the perceived difficulty of the course and the challenge of the course. Moreover, students' ratings of how well the faculty stimulated learning rose more than their ratings of course difficulty. Since recent research (Greenwald, 1996) suggested that faculty who are "easy graders" frequently fare better in student ratings, faculty may have compensated for an anticipated drop in ratings by making their lectures and assignments more intellectually stimulating.

Several post-hoc analyses, however, militated against over-generalizing from this experience. First, the faculty apparently were able to take advantage of a

one-time university-wide grading policy change regarding plus and minus grades that occurred at the same time. The results suggest that this change may have provided department faculty with an easy way to lower the number of A's to meet the new standard without having to lower students' grades by a full letter grade.

The second post-hoc analysis reviewed the maintenance of effects into more recent semesters. These data suggest that some of the desired behavior changes may have proved more difficult to maintain than compliance with the grading policy. In addition, perceived subject matter difficulty may change slowly after the curriculum content and course assignments are revised, compared with perceived instructor challenge which responded to an instructors' performance each semester.

The third post-hoc analysis examined grade changes at a sister institution and suggested that the restraining of high grades is not necessarily rare. However, other evidence strongly suggested unique local effects of the policy. Departmental grades at the neighboring institution had a distinctly different distribution with respect to timing and emphasis. Since the policy focused on reducing the number of A's, it is not surprising that the study institution showed

a steeper drop in A's precisely when the change was instituted.

Nevertheless, the finding is noteworthy. It accords with informal (internet and other) inquiries made by the authors suggesting that both mean increases and decreases are common in the study discipline for institutions in the same state and elsewhere. The finding suggests at least two possible interpretations: 1) faculty may engage in self-monitoring when grades get too far out of line; 2) the policy under study is only one example of the way in which institutions respond to aberrant departments (in terms of grades) within the institution.

The final supplementary analysis revealed that the effects of the policy on grades (and by extension on course challenge) could not be reduced to changes in the two cohorts providing the data. Enrollees' g.p.a.'s inside and outside the department were significantly more similar in the later period. This was further evidence that the department's grading practices came more into line with university-wide norms. The analysis also showed that the clearest evidence of policy effects on students lies in the middle range of general academic achievement.

In sum, the findings of behavior change (stricter grading) were accompanied by some evidence of intended

attitude change (i.e., enhanced scholarly rigor or intellectual challenge and difficulty). These results suggested that the conditions for attitude change alluded to above were met, influencing instructors' rigor and students' perceptions. These results were interesting since, if the perceived demands upon grading exceeded the threshold to be perceived as overly coercive, the result would be outward compliance coupled with a "boomerang effect" towards a more negative attitude (Worchel & Brehm, 1970; Mail, 1993). Such an effect would be unremarkable in a university, a typical normative organization in which compliance depends on internalized directives and coercion tends to foster alienation (Etzioni, 1961).

The study's interest also derived from a controversy underlying the policy. It is hardly unusual for academicians in any field to take issue with an attempt to monitor and influence their grading and pedagogy. The issue was obvious in the adversarial posture taken by some departmental faculty before, during, and after the policy took effect. These facts make it even more curious that the policy enjoyed some success. Basinger (1997) stated that to treat higher grades as a direct result of deficient standards is simplistic, and to require lower grades is to treat symptoms rather than causes or fundamental issues.

Nevertheless, the results of this study suggest that professional faculty will under certain conditions set aside their preferences and grading philosophies in order to meet administrative expectations.

The authors do not suggest that important underlying issues be ignored. The department faculty certainly have not scuttled deeply held convictions about the educational purpose of grading.[6] In some disciplines including the one under study, faculty often have a criterion-referenced orientation to grades which favors a mastery model of student learning (Block, 1971; Geisinger & Rabinowitz, 1979; Hambleton & Murray, 1977). Grades are significant indicators of progress that can be used to prompt students' efforts. In addition, instructors provide critical information when they use grades formatively to correct student errors and shape progress. External pressures to control grade distributions can undermine the process of using grades formatively. Higher levels of reward would become unavailable to many students who required more trials to reach a specified criterion.

A partial resolution of the controversy would rely on increasing the use and importance of externally validated assessment data in program reviews. Such data would provide credible answers to issues regarding the level of student achievement and learning. In

mastery learning terms, mean grades may rise if that rise accompanies a demonstrated increase in student learning. Without such a demonstration, rising grades will evoke the credible evidence that average rises frequently are associated with actual learning declines (Wingspread Group, 1993; Stone, 1995). The demonstration of learning gains accompanying higher grades, however, necessarily avoids the charge of grade inflation, defined as a grade rise without increased achievement (Bejar & Blew, 1981).[7]

This article presented evidence suggesting that respectably high levels of intellectual challenge can be coupled with a policy of restraining the average rise of grades. Thus, maintaining grade standards is not necessarily associated with lower ratings of instruction or "dumbing down" courses. However, such a policy can place faculty in an awkward position when underlying issues regarding disciplinary differences are not fully aired. The meaning of grades may be strongly influenced by disciplinary and subject matter differences or by characteristics of students attracted to different departments (Ekstrom & Villegas, 1992; McKenzie & Tullock, 1981; Summerville, Ridley, & Maris, 1990). However, as long as one system, a 4-level grading scale, saddles all the disciplines with a uniform method of recording progress, awkward

accommodations of the type studied in this paper can be expected.

References

Basinger, D. (1997). Fighting Grade Inflation: A Misguided Effort. *College Teaching, 45*, 88-91.

Bejar, I. I. & Blew., E. O. (1981). Grade inflation and the validity of the scholastic aptitude test. *American Educational Research Journal, 18*, 143-56.

Block, J. H. (Ed.). (1971). *Mastery learning: Theory and practice*. New York: Holt, Rinehart, and Winston.

Cialdini, R. B. (1993). *Influence: The psychology of persuasion*. New York: Quill William Morrow.

Ekstrom, R. B. & Villegas, A. M. (1992). College grades: An exploratory study of policies and practices. *Final Report*. Princeton, N.J.: Educational Testing Service.

Etzioni, A. (1961). *A comparative analysis of complex organizations: On power, involvement, and their correlates*. New York: The Free Press.

Geisinger, K. F. & Rabinowitz, W. (1979). Grading attitudes and practices among college faculty members. In *A spotlight on educational problems*, H. Dahl, A. Lyshe, & P. Rand. (Eds.). Oslo, Norway: Universitesforlaaget, 92-103.

Greenwald, A. G. (1996). Applying social psychology to reveal a major (but correctable) flaw in student evaluations of teaching. ERIC Document

Reproduction Service No. ED 400 754.

Hambleton, R. K. & Murray, J. (1977). A comparative study of faculty and student attitudes toward a variety of college grading purposes and practices. Paper presented at the annual meeting of the National Council on Measurement in Education, ERIC Document Reproduction Service No. ED 141 381.

Mail, P. R. & Van Leeuwen, M. D. (1993). An analysis and restructuring of the diamond model of social response. *Personality and Social Psychology Bulletin, 19*, 106-16.

McKenzie, R. B. & Tullock, G. (1981). *The new world economics*. Homewood, Illinois: Richard D. Erwin.

Rotfeld, H. (1997). Of grades and standards. *Academy of Marketing Science News, 18*, 9.

Stone, J. E. (1995). Inflated grades, inflated enrollments, and inflated budgets: An analysis and call for review at the state level. *Education Policy Analysis Archives, 3*(11).

Summerville, R. M., Ridley, D. R., & Maris, T. L. (1990). Grade inflation in the eighties: The case of urban colleges and universities. *College Teaching, 38*, 33-38.

Winer, B. J. (1971). *Statistical principals in experimental design*. (2nd Edition). New York: McGraw-Hill.

Wingspread Group on Higher Education. (1993). *An*

American imperative: Higher expectations for higher education, an open letter to those concerned about the American future. Racine, Wisconsin: The Johnson Foundation.

Wood, A. L., Ridley, D. R., Summerville, R. M. (1998). University grade inflation through twenty years: An analytical and empirical investigation. Unpublished manuscript.

Worchel, S. & Brehm, J. W. (1970). Effects of threats to attitudinal freedom as a function of agreement with the communicator. *Journal of Personality and Social Psychology, 14,* 18-22.

Table 1. -- Descriptive Statistics and F-values Testing for Interactions Between Discipline and Time (Note: significance probability levels in parentheses; n.s.= not significant)

VARIABLES:	Pre-Policy	Post-Policy	Mn_2-Mn_1	Interaction F-Values
1. Earned Grades				
Dept.: Mn	3.52	3.38	-0.14	
SD	0.75	0.70		
N	2,863	3,063		21.69
---	---	---	---	(p<.005)
Other: Mn	2.67	2.66	-0.01	
SD	1.17	1.19		
N	59,261	57,555		
2. Expected Grades				
Dept.: Mn	3.68	3.46	-0.22	
SD	0.52	0.57		
N	1429	1268		119.53
---	---	---	---	(p<.005)
Other: Mn	3.07	3.18	+0.11	
SD	0.79	0.78		
N	38,612	36,017		
3. Subject Difficulty				
Dept.: Mn	3.08	3.32	+0.24	
SD	1.26	1.23		
N	1464	1264		60.66
---	---	---	---	(p<.005)
Other: Mn	3.45	3.31	-0.14	
SD	1.23	1.28		
N	39,578	35,901		
4. Stimulation				
Dept.: Mn	4.01	4.40	+0.39	
SD	1.29	0.98		
N	508	1005		35.70
---	---	---	---	(p<.005)
Other: Mn	4.16	4.21	+0.05	
SD	1.09	1.08		
N	18,483	31,490		

5. Grading Fairness					
Dept.:	Mn	4.19	4.28	+0.09	
	SD	1.14	1.07		
	N	500	1304		1.98
---		---	---	---	(n.s.)
Other:	Mn	4.40	4.42	+0.02	
	SD	0.96	0.97		
	N	18,416	37,227		
6. Test Content					
Dept.:	Mn	4.17	4.28	+0.11	
	SD	1.06	0.98		
	N	496	1312		1.89
---		---	---	---	(n.s.)
Other:	Mn	4.30	4.34	+0.04	
	SD	0.99	0.97		
	N	18,310	36,940		

Table 2. -- Pre-policy to Post-policy Suggested Re-distribution of A-grades Shown as Percents of all Letter Grades

Unit(s)	Periods	<------- Highest Range		------->	Sums	Lower Range
		A	A-	B+	(A - B+)	B and Below
Study Dept.	Pre-policy	64.3	--	--	64.3	36.7
	Post-policy	36.1	15.8	13.8	65.7	34.3
		A	A-		(A - A-)	B+ and Below
Other Depts.	Pre-policy	27.8	--		27.8	72.2
	Post-policy	20.8	9.2		30.0	70.0

Table 3. -- Descriptive Statistics Related to Discipline and Time for the Neighboring University (Note: Consult Table 1, row 1 to compare these means and mean differences with those of the targeted university)

VARIABLES:	Pre-Policy	Post-Policy	Mn_2-Mn_1	Interaction F-Values
Earned grades				
Dept.: Mn	3.38	3.24	-0.14	
SD	0.89	0.99		
N	26,034	22,831		220.19
---	---	---	---	(p <.005)
Other: Mn	2.60	2.63	+0.03	
SD	1.12	1.18		
N	152,099	138,334		

Notes

1. The authors acknowledge the kind assistance of Frank Dunn and Martha Smith Sharpe of Old Dominion University, Douglas Gallaer and Anda Wood of Christopher Newport University. Thanks are expressed for the helpful suggestions of Robert C. Birney, Chip Byrd (Virginia Commonwealth University), Robert F. Grose, Stephen A. Sivo (James Madison University) and Clinton B. Walker. Requests for reprints should be sent to the first author at the Office of Institutional Effectiveness, Christopher Newport University, 50 Shoe Lane, Newport News, Virginia 23606-2988.
2. This article deliberately focused upon an underlying issue, the empirical distinction between grading difficulty and course challenge, rather than a specific discipline or department issue. Therefore, to clarify the general problem while protecting the anonymity of any department or its members, the article avoided reference to the specific department providing much of the data. Its contribution to this article, albeit anonymous, also is gratefully acknowledged.
3. The authors are indebted to Chip Byrd for the suggestion leading to the analysis described in this paragraph.
4. Although analysis of variance theoretically rests on the assumption of homogeneity of population variances, it is robust with respect to all but fairly large deviations from that assumption. Screening done based on the recommendations of Winer (1971, p. 206) supported

the validity of analysis of variance for this study.

5. To put changes in perspective one needs to remember that the student ratings of instruction are made in discrete intervals (e.g. 1, 2, 3, 4, and 5) rather than on a continuous scale. Thus each change of 0.1 in a mean rating is the equivalent of having 10 percent of each class raise or lower a rating by a full point. A mean change of 0.5 would be the equivalent of changing the opinion of half the students.
6. When the grading policy began, department members objected on a variety of grounds, brought the issue before the faculty governing body, and threatened appeal to the AAUP. For its next program review five years later, the department carefully presented data to support its position, including outside review of syllabi suggesting good representation of academically challenging assignments.
7. More generally, grade inflation is decline in the value of grades in the coin of student achievement. Thus, theoretically grade inflation is possible even if mean grades are level or declining. (Reference: Wood, Ridley, & Summerville, 1998).